SOLUTION OVERVIEW

Image courtesy of Kinetic Vision

REAL-TIME AI PERFORMANCE AT THE EDGE

AI Is Now Everywhere

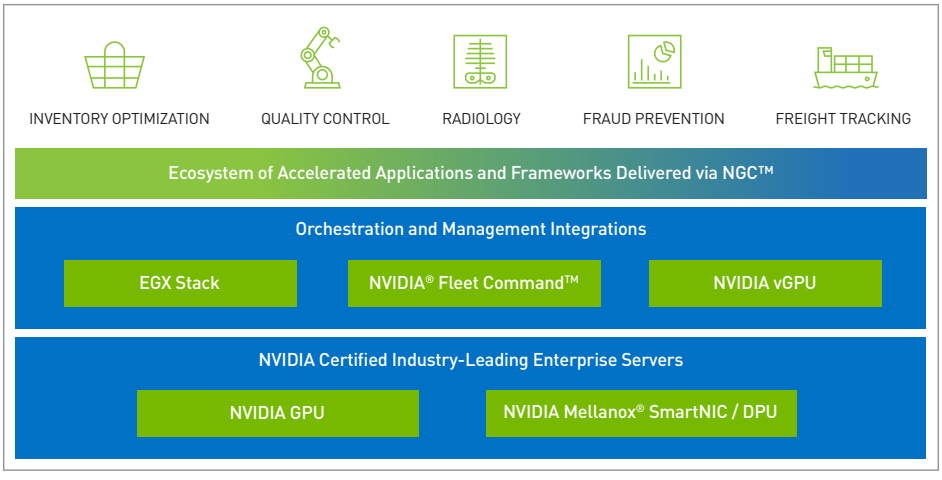
As businesses across industries grapple with vast amounts of data, more complex operations, and more dynamic markets, edge AI is playing a growing role in helping them rapidly respond. Through a combination of computing power, AI technology, data analytics, and advanced connectivity, the edge extends compute capabilities from data centers out to the edge of networks, allowing organizations to act quickly on data where it's captured. Reducing the distance between where data is captured and where it's processed not only alleviates data transit costs, but also improves latency, bandwidth utilization, and infrastructure costs.

Addressing the Requirements of Edge AI

AI at the edge comes with a unique set of requirements. Edge systems, dispersed across vast physical distances, lack the centrality that a data center presents. Software or system updates either need to be performed manually or need to be centrally managed to easily deploy, manage, and scale software across vast fleets of devices. Moreover, the security requirements for edge computing infrastructure differ from cloud or data center computing models. Edge locations lack the physical security that data centers have, so an end-to-end security model that protects both the application IP and the sensor data is paramount for a successful deployment.

NVIDIA EGX Platform for Accelerated Edge AI

The NVIDIA EGX platform allows enterprise IT to deliver diverse applications on high-performance and cost-effective infrastructure. The platform is a combination of high-performance GPU computing and high-speed, secure networking in NVIDIA-



The EGX platform supports a vast suite of accelerated applications for edge AI, delivering faster insights where they matter the most.

Generate faster insights from AI deployed across thousands of devices using the NVIDIA EGX™ platform.

KEY USE CASES

- Inventory optimization and shrinkage reduction in retail stores
- Automation and quality control in manufacturing facilities
- Radiology and patient care in hospitals
- 5G multi-access edge computing (MEC) and virtual radio area network (vRAN) for telecommunications providers
- Fraud prevention and recommendations for financial institutions
- Freight tracking and route optimization for efficient logistics

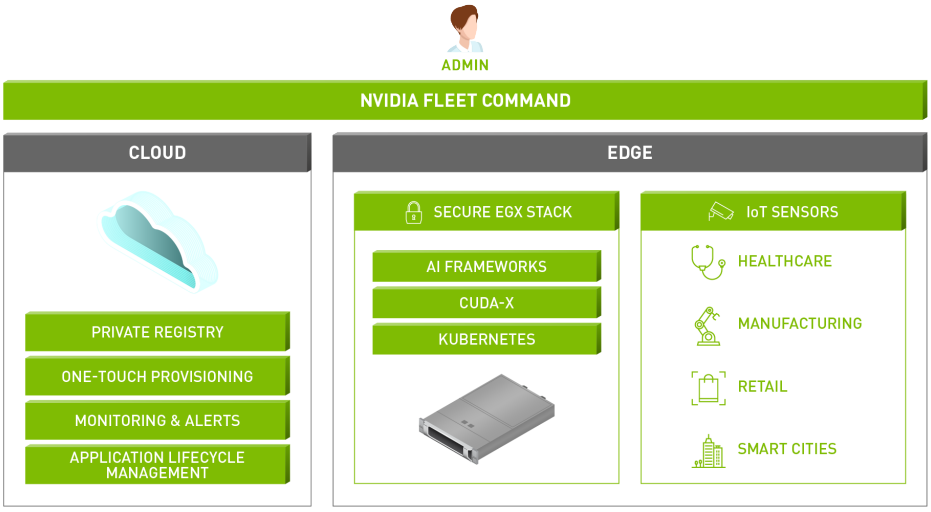
PROOF POINTS

- Customers improved real-time fraud detection by 10 percent, lowered their server capacity by nearly 8X, and deployed a new fraud detection service that operated worldwide 24/7 in real time to protect customer transactions from potential fraud. A requirement that CPU-only servers couldn't meet.

Certified Systems™, built and sold by our partners. The EGX platform allows customers to prepare for the future while driving down costs by standardizing on a single unified architecture for easy management, deployment, operation, and monitoring. The EGX platform supports a vast suite of accelerated applications for edge AI, delivering faster insights where they matter the most.

Securely Deploy, Manage, and Scale Applications with NVIDIA Fleet Command

Fleet Command is a hybrid-cloud platform for managing and scaling AI deployments across dozens or up to millions of servers or edge devices. Fleet Command allows IT departments to securely and remotely manage a large-scale fleet of deployed systems. Instead of spending weeks planning and executing deployment plans, in minutes, administrators can bring AI to networks of retail stores, warehouses, hospitals, or city streets. Administrators can add or delete applications, update system software over the air, and monitor the health of devices spread across vast distances from a single control plane.



Securing the Edge

In addition to accelerated computing and simplified deployments, NVIDIA solutions for edge computing offer industry-leading security protocols to ensure data is always protected. All processed data is encrypted in transit and at rest and secure and measured boot protects the AI runtime from being tampered with. Because systems are on premises to process local sensor feeds, organizations maintain where sensor data is stored. Furthermore, AI applications deployed using Fleet Command are scanned for vulnerabilities and malware and offer signed containers, certifying that every application deployed is secure.

NVIDIA-CERTIFIED SYSTEMS

- Confidently deploy scalable hardware and software solutions that securely and optimally run accelerated workloads.

NVIDIA CERTIFIED

NVIDIA